

1 TCCGGGGTCC GCACCGGGCC TGAGTCGGTC CGAGGGCGTC CCAGGGAGCAG  
51 CTGCCCCTGC GGAACAGCAC TATGGGCTTC TCTTCTGAGC TGTGCAGCCC  
101 CCAGGGCCAC GGGGTCTGC AGCAAATGCA GGAGGGCCGAG CTTCGTCTAC  
151 TGGAGGGCAT GAGAAAGTGG ATGGCCAGC GGGTCAAGAG TGACAGGGAG  
201 TATGCAGGAC TGCTTCACCA CATGTCCCTG CAGGACAGTG GGGGCCAGAG  
251 CCGGGCCATC AGCCCTGACA GCCCCATCAG TCAGTCCTGG GCTGAGATCA  
301 CCAGCCAAAC TGAGGGCCTG AGCCGCTTGC TGCCGCAGCA CGCAGAGGAT  
351 CTGAACTCAG GGCCCCTGAG CAAGCTGAGC CTGCTCATCC GGGAACGGCA  
401 GCAGCTTCGC AAGACCTACA GCGAGCAGTG GCAGCAGCTG CAGCAGGGAGC  
451 TCACCAAGAC CCACAGCCAG GACATTGAGA AGCTGAAGAG CCAGTACCGA  
501 GCTCTGGCAC GGGACAGTGC CCAAGCCAAG CGCAAGTACC AGGAGGCCAG  
551 CAAAGACAAG GACCGTGACA AGGCCAAGGA CAAGTATGTG CGCAGCCTGT  
601 GGAAGCTCTT TGCTCACCAC AACCCTATG TGCTGGGCGT GCGGGCTGCG  
651 CAGCTACACC ACCAGCACCA CCACCAGCTC CTGCTGCCCG GCCTGCTGCG  
701 GTCACTGCAG GACCTGCACG AGGAGATGGC TTGCATCCTG AAGGAGATCC  
751 TGCAGGAATA CCTGGAGATT AGCAGCCTGG TGCAAGGATGA GGTGGTGGCC  
801 ATTCAACCGGG AGATGGCTGC AGCTGCTGCC CGCATCCAGC CTGAGGGCTGA  
851 GTACCAAGGC TTCCCTGCGAC AGTATGGGTC CGCACCTGAC GTCCCACCC  
901 GTGTCACGTT CGATGAGTCA CTGCTTGAGG AGGGTGAACC GCTGGAGCCT  
951 GGGGAGCTCC AGCTGAACGA GCTGACTGTG GAGAGCGTGC AGCACACGCT  
1001 GACCTCAGTG ACAGATGAGC TGGCTGTGGC CACCGAGATG GTGTTCAGGC  
1051 GGCAGGAGAT GGTTACCGCAG CTGCAACAGG AGCTCCGGAA TGAAGAGGAG  
1101 AACACCCACC CCCGGAGCG GGTGCAGCTG CTGGGCAAGA GGCAAGTGCT  
1151 GCAAGAAGCA CTGCAGGGGC TGCAAGGTAGC GCTGTGCAGC CAGGCCAAC  
1201 TGCAGGCCCA GCAGGAGTTG CTGCAGACCA AGCTGGAGCA CCTGGGCC  
1251 GGCAGGCCCG CGCCTGTGCT GTCCTGCAG GATGACCGCC ACTCCACGTC  
1301 GTCCTCGGAG CAGGAGCGAG AGGGGGGAAG GACACCCACG CTGGAGATCC  
1351 TTAAGAGCCA CATCTCAGGA ATCTTCCGCC CCAAGTTCTC GAACCTGTAC  
1401 CGACTGGAAG GGGAAAGGCTT TCCTAGCATT CCTTTGCTCA TCGACCAACCT  
1451 ACTGAGCACC CAGCAGCCCC TCACCAAGAA GAGTGGTGTG GTCCTGCACA  
1501 GGGCTGTGCC CAAGGACAAG TGGGTGCTGA ACCATGAGGA CCTGGTGTG  
1551 GGTGAGCAGA TTGGACGGGG GAACTTTGGC GAAGTGTCA GCGGACGCCT  
1601 GCGAGCCGAC AACACCCCTGG TGGCGGTGAA GTCTTGTGCA GAGACGCTCC  
1651 CACCTGACCT CAAGGCCAAG TTTCTACAGG AAGCGAGGAT CCTGAAGCAG  
1701 TACAGCCACC CCAACATCGT GCGTCTCATT GGTGTCTGCA CCCAGAAGCA  
1751 GCCCATCTAC ATCGTCATGG AGCTTGTGCA GGGGGCGAC TTCCCTGACCT  
1801 TCCTCCGCAC GGAGGGGGCC CGCCTGCAGG TGAAAGACTCT GCTGCAGATG  
1851 GTGGGGGATG CAGCTGCTGG CATGGAGTAC CTGGAGAGCA AGTGCTGCAT  
1901 CCACCGGGAC CTGGCTGCTC GGAACCTGCCT GGTGACAGAG AAGAATGTCC  
1951 TGAAGATCAG TGACTTTGGG ATGTCCCGAG AGGAAGCCGA TGGGGTCTAT  
2001 GCAGCCTCAG GGGGCCTCAG ACAAGTCCCC GTGAAGTGGG CCGCACCTGA  
2051 GGCCTTAAC TACGGCCGCT ACTCCTCCGA AAGCGACGTG TGGAGCTTG  
2101 GCATCTTGCT CTGGGAGACC TTCAGCCTGG GGGCCTCCCC CTATCCCAAC  
2151 CTCAGCAATC AGCAGACACG GGAGTTGTG GAGAAGGGGG GCGTCTGCC  
2201 CTGCCCCAGAG CTGTGTCTG ATGCCGTGTT CAGGCTCATG GAGCAGTGCT  
2251 GGGCCTATGA GCCTGGGCAG CGGCCAGCT TCAGCACCCT ATACCAGGAG

FIG. 1A

2301 CTGCAGAGCA TCCGAAAGCG GCATCGGTGA GGCTGGGACC CCCTTCTCAA  
2351 GCTGGTGGCC TCTGCAGGCC TAGGTGCAGC TCCTCAGCGG CTCCAGCTCA  
2401 TATGCTGACA GCTCTTCACA GTCCTGGACT CCTGCCACCA GCATCCACAC  
2451 TGCCGGCAGG ATGCAGCGCC GTGTCCTCTC TGTGTCCCTG CTGCTGCCAG  
2501 GGCTTCCTCT TCCGGCAGA AACAAATAAAA CCACTTGTGC CCACTGAAAAA  
2551 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA  
2601 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA  
2651 AAAAAAAAAA AAAAAAAAAA AAAA (SEQ ID NO:1)

**FEATURES:**

5'UTR: 1-71  
Start Codon: 72  
Stop Codon: 2328  
3'UTR: 2331

**Homologous proteins:**

Top 10 BLAST Hits

|  | Score | E     |
|--|-------|-------|
| CRA 18000004928524 /altid=gi 4503687 /def=ref NP_001996.1  feli... | 1364  | 0.0   |
| CRA 18000004928925 /altid=gi 400127 /def=sp P07332 FES_HUMAN PR... | 1361  | 0.0   |
| CRA 335001098689057 /altid=gi 11433086 /def=ref XP_007718.1  fe... | 1361  | 0.0   |
| CRA 18000004944482 /altid=gi 66835 /def=pir TVCTFF protein-tyr...  | 1285  | 0.0   |
| CRA 18000004883448 /altid=gi 1345986 /def=sp P14238 FES_FELCA P... | 1285  | 0.0   |
| CRA 18000004938794 /altid=gi 125356 /def=sp P16879 FES_MOUSE PR... | 1249  | 0.0   |
| CRA 18000004944484 /altid=gi 125354 /def=sp P00543 FES_FSVST TY... | 647   | 0.0   |
| CRA 18000004958077 /altid=gi 323873 /def=gb AAA43041.1  (J02087... | 621   | e-176 |
| CRA 18000004944483 /altid=gi 125353 /def=sp P00542 FES_FSVGA TY... | 621   | e-176 |
| CRA 108000000500738 /altid=gi 7548235 /def=gb AAA43046.2  (J020... | 603   | e-171 |

BLAST dbEST hits:

|   | Score | E   |
|---|-------|-----|
| gi 12875454 /dataset=dbest /taxon=960...  | 1669  | 0.0 |
| gi 12259598 /dataset=dbest /taxon=960...  | 924   | 0.0 |
| gi 5526793 /dataset=dbest /taxon=9606 ... | 856   | 0.0 |
| gi 1501859 /dataset=dbest /taxon=9606 ... | 722   | 0.0 |
| gi 9097978 /dataset=dbest /taxon=9606...  | 714   | 0.0 |
| gi 6131861 /dataset=dbest /taxon=9606 ... | 682   | 0.0 |
| gi 12447497 /dataset=dbest /taxon=96...   | 674   | 0.0 |

EXPRESSION INFORMATION FOR MODULATORY USE:

Library source:

From BLAST dbEST hits:

gi|12875454 Placenta  
gi|12259598 Lung-tumor  
gi|5526793 Kidney 2 pooled tumors  
gi|1501859 Pregnant uterus  
gi|9097978 Pediatric pre-B cell lymphoblastic leukemia  
gi|6131861 Stomach poorly differentiated adenocarcinoma with signet ring cells  
gi|12447497 Placenta normal

From tissue screening panels:

Hippocampus

FIG. 1C

1 MGFSELCS P QGHGVLQQMQ EAELRLLEG M RKWMAQRVKS DREYAGLLHH  
51 MSLQDSGGQS RAISPDSPIS QSWAEITSQT EGLSRLLRQH AEDLNSGPLS  
101 KLSLLIRERQ QLRKTYSEQW QQLQQELTKT HSQDIEKLKS QYRALARDSA  
151 QAKRKYQEAS KDKDRDKAKD KYVRSLWKL F AHHNRYVLGV RAAQLHHQHH  
201 HQLLLPGLLR SLQDLHEEMA CILKEILQYE LEISSLVQDE VVAIHREMAA  
251 AAARIQPEAE YQGFLRQYGS APDVPPCVTF DESLLEEGER LEPGEQLNE  
301 LTVESVQHTL TSVTDELAVA TEMVFRQQEM VTQLQQELRN EEENTHPRER  
351 VQLLGKRQVL QEALQGLQVA LCSQAKLQAQ QELLQTKLEH LGPGEPPPVL  
401 LLQDDRHSTS SSEQEREGGR TPTLEILKSH ISGIFRPKFS NLYRLEGE GF  
451 PSIPLLIDHIL LSTQQPLTKK SGWLHRAVP KDKWLNHED LVLGEQIGRG  
501 NFGEVFSGRL RADNTLVAVK SCRETLPPDL KAKFLQEARI LKQYSHPNIV  
551 RLIGVCTQKQ PIYIVMELVQ GGDFLTFLRT EGARLRVKT LQMVGDAAG  
601 MEYLESKCCI HRDLAARNCL VTEKNVLKIS DFGMSREEAD GVYAAASGGLR  
651 QVPVKWTAPE ALNYGRYSSE SDWWSFGILL WETFSLGASP YPNLSNQQTR  
701 EFVEKGGRLP CPELCPDAVF RLMEQCWAYE PGQRPSFSTI YQELQSIRKR  
751 HR (SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

[1] PDO000001 PS00001 ASN\_GLYCOSYLATION

N-glycosylation site

693-696 NLSN

[2] PDO000005 PS00005 PKC\_PHOSPHO\_SITE

Protein kinase C phosphorylation site

Number of matches: 7

|   |         |     |
|---|---------|-----|
| 1 | 40-42   | SDR |
| 2 | 468-470 | TKK |
| 3 | 507-509 | SGR |
| 4 | 521-523 | SCR |
| 5 | 557-559 | TQK |
| 6 | 622-624 | TEK |
| 7 | 746-748 | SIR |

[3] PDO000006 PS00006 CK2\_PHOSPHO\_SITE

Casein kinase II phosphorylation site

FIG. 2A

Number of matches: 17

- 1 40-43 SDRE
- 2 52-55 SLQD
- 3 211-214 SLQD
- 4 72-75 SWAE
- 5 78-81 SQTE
- 6 115-118 TYSE
- 7 52-55 SLQD
- 8 211-214 SLQD
- 9 270-273 SAPD
- 10 279-282 TFDE
- 11 283-286 SLLE
- 12 312-315 SVTD
- 13 386-389 TKLE
- 14 410-413 SSSE
- 15 412-415 SEQE
- 16 521-524 SCRE
- 17 635-638 SREE

[4] PDOC00007 PS00007 TYR\_PHOSPHO\_SITE  
Tyrosine kinase phosphorylation site

Number of matches: 3

- 1 37-44 RVKSDREY
- 2 254-261 RIQPEAEY
- 3 636-643 REEADGVY

[5] PDOC00008 PS00008 MYRISTYL  
N-myristoylation site

Number of matches: 6

- 1 58-63 GQSRAI
- 2 189-194 GVRAAQ
- 3 207-212 GLLRSL
- 4 366-371 GLQVAL
- 5 641-646 GVYAAS
- 6 732-737 GQRPSF

[6] PDOC00009 PS00009 AMIDATION  
Amidation site

354-357 LGKR

FIG. 2B

[7] PDOC00100 PS00107 PROTEIN\_KINASE\_ATP  
Protein kinases ATP-binding region signature

497-520 IGRGNFGEVFSGRLRADNTLVAVK

[8] PDOC00100 PS00109 PROTEIN\_KINASE\_TYR  
Tyrosine protein kinases specific active-site signature

609-621 CIHRDLAARNCLV

Membrane spanning structure and domains:

| Helix | Begin | End | Score | Certainty |
|-------|-------|-----|-------|-----------|
| 1     | 672   | 692 | 1.129 | Certain   |

BLAST Alignment to Top Hit:

>CRA|18000004928524 /altid=gi|4503687 /def=ref|NP\_001996.1| feline  
sarcoma (Snyder-Theilen) viral (v-fes)/Fujinami avian  
sarcoma (PRCII) viral (v-fps) oncogene homolog; Oncogene  
FES, feline sarcoma virus [Homo sapiens] /org=Homo  
sapiens /taxon=9606 /dataset=nraa /length=822  
Length = 822

Score = 1364 bits (3491), Expect = 0.0

Identities = 706/822 (85%), Positives = 716/822 (86%), Gaps = 70/822 (8%)

Frame = +3

Query: 72 MGFSELCSPOGHVLQQMQEAEIQLLEGMRKWMQRVKSDREYAGLLHHMSLQDSGGQS 251

MGFSELCSPOGHVLQQMQEAEIQLLEGMRKWMQRVKSDREYAGLLHHMSLQDSGGQS

Sbjct: 1 MGFSELCSPOGHVLQQMQEAEIQLLEGMRKWMQRVKSDREYAGLLHHMSLQDSGGQS 60

Query: 252 RAISPDSPISQSWAEITSQTEGLSRLLRQHAEDLNSGPLSKLSLLIRERQQLRKTYSEQW 431

RAISPDSPISQSWAEITSQTEGLSRLLRQHAEDLNSGPLSKLSLLIRERQQLRKTYSEQW

Sbjct: 61 RAISPDSPISQSWAEITSQTEGLSRLLRQHAEDLNSGPLSKLSLLIRERQQLRKTYSEQW 120

Query: 432 QQLQQELTKTHSQDIEKLKSQYRALARDSAQAKRKYQEASKDKDRDKAKDKYVRSLWLF 611

QQLQQELTKTHSQDIEKLKSQYRALARDSAQAKRKYQEASKDKDRDKAKDKYVRSLWLF

Sbjct: 121 QQLQQELTKTHSQDIEKLKSQYRALARDSAQAKRKYQEASKDKDRDKAKDKYVRSLWLF 180

Query: 612 AHHNRYVLGVRAAQLHHQHHHQLLPGLLRLQLHEEMACILKEILQEYLEISSLVQDE 791

AHHNRYVLGVRAAQLHHQHHHQLLPGLLRLQLHEEMACILKEILQEYLEISSLVQDE

Sbjct: 181 AHHNRYVLGVRAAQLHHQHHHQLLPGLLRLQLHEEMACILKEILQEYLEISSLVQDE 240

Query: 792 WAIHREMAAAARIQPEAEYQGFLRQYGSAPDVPPCVTFDESLLEEGEPLEPGELQLNE 971

WAIHREMAAAARIQPEAEYQGFLRQYGSAPDVPPCVTFDESLLEEGEPLEPGELQLNE

Sbjct: 241 WAIHREMAAAARIQPEAEYQGFLRQYGSAPDVPPCVTFDESLLEEGEPLEPGELQLNE 300

FIG. 2C

Query: 972 LTVESVQHTLTSVTDELAVATEMFRRQEMVTQLQQELRNEEENTHPRERVQLLGKRQL 1151  
 LTVESVQHTLTSVTDELAVATEMFRRQEMVTQLQQELRNEEENTHPRERVQLLGKRQL  
 Sbjct: 301 LTVESVQHTLTSVTDELAVATEMFRRQEMVTQLQQELRNEEENTHPRERVQLLGKRQL 360

Query: 1152 QEAQGLQVALCSQAK-----LQAQQE-----LLQTKLEHLGPGE----- 1256  
 QEAQGLQVALCSQAK LQ + E LLQ E  
 Sbjct: 361 QEAQGLQVALCSQAKLQAQQELLQTKLEHLGPGEPPPVLQLQDDRHSTSSEQEREGGR 420

Query: 1257 -----PPPVLQLQDDR-----HSTSS 1304  
 PPP+ L+ + + HS  
 Sbjct: 421 TPTLEILKSHISGIFRPKFLPPLQLIPEVQKPLHEQLWYHGAIPRAEVAELLVHSQDF 480

Query: 1305 SEQEREGGRTPTLEILKSHISGIFR-PKFSNLYRLEGEGLPSIPLLIDHLLSTQQPLTKK 1481  
 +E +G + L +L + F NLYRLEGEGLPSIPLLIDHLLSTQQPLTKK  
 Sbjct: 481 LVRESQGKQEYVLSVLWDGLPRHFIIQSLDNLYRLEGEGLPSIPLLIDHLLSTQQPLTKK 540

Query: 1482 SGWLHRAVPKDKWLNHEDLVLGEQIGRGNFGEVFSGRLRADNTLVAVKSCRETLPPDL 1661  
 SGWLHRAVPKDKWLNHEDLVLGEQIGRGNFGEVFSGRLRADNTLVAVKSCRETLPPDL  
 Sbjct: 541 SGWLHRAVPKDKWLNHEDLVLGEQIGRGNFGEVFSGRLRADNTLVAVKSCRETLPPDL 600

Query: 1662 KAKFLQEARILKQYSHPNIVRLIGVCTQKQPIYIVMELVQGGDFLTFLRTEGARLRVKTL 1841  
 KAKFLQEARILKQYSHPNIVRLIGVCTQKQPIYIVMELVQGGDFLTFLRTEGARLRVKTL  
 Sbjct: 601 KAKFLQEARILKQYSHPNIVRLIGVCTQKQPIYIVMELVQGGDFLTFLRTEGARLRVKTL 660

Query: 1842 LQMVGDAAAGMEYLESKCCIHRDLAARNCLVTEKNVLKISDFGMSREEADGVYAAASGLR 2021  
 LQMVGDAAAGMEYLESKCCIHRDLAARNCLVTEKNVLKISDFGMSREEADGVYAAASGLR  
 Sbjct: 661 LQMVGDAAAGMEYLESKCCIHRDLAARNCLVTEKNVLKISDFGMSREEADGVYAAASGLR 720

Query: 2022 QVPVKWTAPEALNYGRYSSESDWSFGILLWETFSLGASPYPNLSNQQTREFVEKGRLP 2201  
 QVPVKWTAPEALNYGRYSSESDWSFGILLWETFSLGASPYPNLSNQQTREFVEKGRLP  
 Sbjct: 721 QVPVKWTAPEALNYGRYSSESDWSFGILLWETFSLGASPYPNLSNQQTREFVEKGRLP 780

Query: 2202 CPELCPDAVFRLMEQCWAYEPGQRPSFSTIYQELQSIRKRHR 2327  
 CPELCPDAVFRLMEQCWAYEPGQRPSFSTIYQELQSIRKRHR  
 Sbjct: 781 CPELCPDAVFRLMEQCWAYEPGQRPSFSTIYQELQSIRKRHR 822 (SEQ ID NO:4)

Hammer search results (Pfam):

| Model   | Description                      | Score | E-value | N |
|---------|----------------------------------|-------|---------|---|
| CE00287 | CE00287 PTK_Eph_orphan_receptor  | 285.3 | 7.8e-82 | 1 |
| PF00069 | Eukaryotic protein kinase domain | 271.5 | 1.1e-77 | 1 |
| CE00292 | CE00292 PTK_membrane_span        | 240.7 | 2.1e-68 | 1 |
| CE00290 | CE00290 PTK_Trk_family           | 232.7 | 5.4e-66 | 1 |
| CE00291 | CE00291 PTK_fgf_receptor         | 190.3 | 3e-53   | 1 |
| CE00031 | CE00031 VEGFR                    | 149.8 | 6.1e-45 | 2 |

FIG. 2D

|         |  |        |         |   |
|---------|--|--------|---------|---|
| PF00611 | Fes/CIP4 homology domain                   | 149.0  | 1.9e-42 | 1 |
| CE00334 | E00334 urotrophin_receptor                 | 143.6  | 3.5e-45 | 1 |
| CE00286 | E00286 PTK_EGF_receptor                    | 141.3  | 1.7e-38 | 1 |
| CE00204 | CE00204 FIBROBLAST_GROWTH_RECECTOR         | 124.9  | 8.1e-34 | 2 |
| CE00288 | CE00288 PTK_Insulin_receptor               | 110.9  | 2.5e-29 | 1 |
| CE00202 | CE00202 EPHRIN_TYPE_A_RECECTOR             | 108.3  | 1.9e-29 | 3 |
| CE00203 | CE00203 ERBB_RECECTOR                      | 81.8   | 2.1e-22 | 1 |
| CE00549 | CE00549 NGFR                               | 69.2   | 7.4e-20 | 1 |
| CE00201 | CE00201 EPHRIN_TYPE_B_RECECTOR             | 57.0   | 5.9e-15 | 4 |
| CE00359 | E00359 bone_morphogenetic_protein_receptor | 26.3   | 9.4e-07 | 1 |
| PF00017 | Src homology domain 2                      | 10.4   | 0.024   | 1 |
| PF00422 | ATP synthase Alpha chain, C terminal       | 7.7    | 0.11    | 1 |
| CE00289 | CE00289 PTK_PDGF_receptor                  | -33.8  | 8.2e-05 | 1 |
| CE00016 | CE00016 GSK_glycogen_synthase_kinase       | -290.0 | 1.2     | 1 |

Parsed for domains:

| Model   | Domain | seq-f | seq-t | hmm-f | hmm-t        | score  | E-value |
|---------|--------|-------|-------|-------|--------------|--------|---------|
| PF00611 | 1/1    | 1     | 94    | [.    | 1 109 □      | 149.0  | 1.9e-42 |
| PF00422 | 1/1    | 210   | 234   | ..    | 111 135 .]   | 7.7    | 0.11    |
| PF00017 | 1/1    | 441   | 460   | ..    | 60 79 .]     | 10.4   | 0.024   |
| CE00201 | 1/4    | 497   | 510   | ..    | 662 675 ..   | 0.1    | 17      |
| CE00202 | 1/3    | 497   | 511   | ..    | 680 694 ..   | 3.2    | 2.4     |
| CE00202 | 2/3    | 534   | 557   | ..    | 721 744 ..   | 5.3    | 0.63    |
| CE00201 | 2/4    | 534   | 567   | ..    | 702 735 ..   | 4.2    | 1.3     |
| CE00031 | 1/2    | 483   | 580   | ..    | 868 972 ..   | 4.5    | 0.17    |
| CE00289 | 1/1    | 489   | 588   | ..    | 1 109 □      | -33.8  | 8.2e-05 |
| CE00204 | 1/2    | 600   | 636   | ..    | 649 685 ..   | 68.1   | 3.5e-18 |
| CE00201 | 3/4    | 598   | 636   | ..    | 766 804 ..   | 21.3   | 2.9e-05 |
| CE00359 | 1/1    | 495   | 636   | ..    | 145 299 ..   | 26.3   | 9.4e-07 |
| CE00204 | 2/2    | 653   | 692   | ..    | 702 741 ..   | 56.1   | 7.1e-15 |
| CE00201 | 4/4    | 653   | 697   | ..    | 824 868 ..   | 30.5   | 9.5e-08 |
| CE00334 | 1/1    | 490   | 734   | ..    | 539 803 ..   | 143.6  | 3.5e-45 |
| CE00203 | 1/1    | 598   | 737   | ..    | 850 989 ..   | 81.8   | 2.1e-22 |
| CE00202 | 3/3    | 598   | 740   | ..    | 827 970 ..   | 99.5   | 5.2e-27 |
| CE00290 | 1/1    | 491   | 744   | ..    | 1 282 □      | 232.7  | 5.4e-66 |
| CE00291 | 1/1    | 491   | 744   | ..    | 1 285 □      | 190.3  | 3e-53   |
| CE00286 | 1/1    | 491   | 744   | ..    | 1 263 □      | 141.3  | 1.7e-38 |
| CE00292 | 1/1    | 491   | 744   | ..    | 1 288 □      | 240.7  | 2.1e-68 |
| CE00031 | 2/2    | 598   | 744   | ..    | 1056 1203 .. | 145.3  | 1.4e-43 |
| CE00287 | 1/1    | 491   | 744   | ..    | 1 260 □      | 285.3  | 7.8e-82 |
| CE00288 | 1/1    | 491   | 744   | ..    | 1 269 □      | 110.9  | 2.5e-29 |
| CE00549 | 1/1    | 598   | 745   | ..    | 693 840 ..   | 69.2   | 7.4e-20 |
| PF00069 | 1/1    | 491   | 746   | ..    | 1 278 □      | 271.5  | 1.1e-77 |
| CE00016 | 1/1    | 431   | 751   | ..    | 1 433 □      | -290.0 | 1.2     |

FIG. 2E

1 CTGGCCACCA GGCTGGCGCA GCCAAGGCG AAGCTCTGGC TGAACCTGT  
51 GCTGGTGTCC TGACCACCT CCCCTCTCTT GCACCCGCCT CTCCCGTCAG  
101 GGCCCAAGTC CCTGTTTCT GAGCCCGGGC TGCCCTGGCT GTTGGCACTC  
151 ACAGACCTGG AGCCCCGGG TGGGTGGTGG GGAGGGGCGC TGGCCCAGCC  
201 GGCTCTCTG GCCTCCCACC CGATGCTGCT TTCCCTGTG GGGATCTCAG  
251 GGGCTGTTG AGGATATATT TTCACTTTGT GATTATTCA CTTTAGATGC  
301 TGATGATTG TTTTGATT TTAATGGGG GTAGCAGCTG GACTACCCAC  
351 GTTCTCACAC CCACCGTCCG CCCTGCTCCT CCCTGGCTGC CCTGGCCCTG  
401 AGGTGTGGGG GCTGCAGCAT GTTGCTGAGG AGTGAGGAAT AGTTGAGCCC  
451 CAAGTCCTGA AGAGGGGGC CAGCCAGGGC GGCTCAAGGA AAGGGGGTCC  
501 CAGTGGGAGG GGCAGGCTGA CATCTGTGTT TCAAGTGGGG CTCGCCATGC  
551 CGGGGGTTCA TAGGTCACTG GCTCTCCAAG TGCCAGAGGT GGGCAGGTGG  
601 TGGCACTGAG CCCCCCAAC ACTGTGCCCT GGTGGAGAAA GCACTGACCT  
651 GTCACTGCCCT CCTCAAACCT CCTCTTCTGA CGTGCCTTT GCACCCCTCC  
701 CATTAGGACA ATCAGTCCCC TCCCATCTGG GAGTCCCCTT TTCTTTCTA  
751 CCCTAGCCAT TCCTGGTACC CAGCCATCTG CCCAGGGTG CCCCCTCCTC  
801 TCCCATCCCC CTGCCCTCGT GGCCAGCCG GCTGGTTTG TAAGATACTG  
851 GGTGGTGCA CAGTGATT TTTCTTGAA TTTAAACAGG CCCAGCATTG  
901 CTGGTTCTAT TTAATGGACA TGAGATAATG TTAGAGGTTT TAAAGTGATT  
951 AAACGTGCAG ACTATGCAA CCAGGCCAG TCTCCAGTGT GGTACCGTTG  
1001 CTCCCTGCATC GCAGCTGAGG ATAGGGGCC AGTTAGGCCT ACACAGTGGC  
1051 CTGCCTGCCT GGATGTGGGC CCAAGTCAGA AGGCCAAAGT CCTCCAAGGG  
1101 GCGGGAGGAT GCGCCAGCCC CTAGTGGAGG AGCTGGTGC CCGGGGTGG  
1151 GGCTGGTAC CCCTGGTCCT CAGGAGCTGA GCACTAAACT CCCAAAGTCC  
1201 TGTTTCCAG CAGTGTGAAG AACTGGGCC ATTGTGTCTT CCTGGGCTGA  
1251 AGTGTCTGG TCGCCACAGG CTATAGGCT GAGGCCTAAG GTGGAGGGAG  
1301 GCCTGACTGA ATCAAGATGA CTTCTTGTT GGAGCCTGAG TCCCAATGG  
1351 AAAACTCCAC GCCTGTCCGC TCCCCAACCC CTGCCCCCTTG ATTTCCCCAG  
1401 GTCTCCCTTG GGACAGGAAG CCCCTGCCTG GGGTAGGAG GATGGGGACA  
1451 AAACCACTAG GATCTGTATC CGAGAAGCAG TCTCTGTTCG GGATATTAC  
1501 TTGGAAATT TATTCAAATG GAAGCTGGCG CCTGAGCCTC TCCTTAGGGA  
1551 ATTCGGTGAAG GTGGGGAGGG CTGGGACCAAG GTTCCCTCTT TTCTCTCTG  
1601 CGGTGGCCCT GGCCTGGTGC TAGGACTGCG CGCCTCCCT CAGTACCGC  
1651 GGACACCCCTG GGCTTCCCTG GGCCCAAGCAT CTGCTGGGG CCTCGCCCTG  
1701 GGCTCCCCCT CCTGACCCCCC ACCTTGCCTC CCTTCCCGGT GTTCCCGGGG  
1751 CGCTGCCGGG CCCTGGGGCC TGCGGGGGCGC GGGCGGCTCT TGGCTGGGCC  
1801 ATTCTTCCCC GGCCCCCTCC TCCCTTCCGT TTCCGTGGCC GTGCGGCCGG  
1851 CTAGAGGCTG CGGCCCCAGCG CGGAGCAGGG GGGCTGGCAG GCGTCGGGAC  
1901 GGTGGGGCCG GTCCCCGCCG CCCCTTCCCC TCCACAGGCC CGCCCCGGGG  
1951 CCTGGGCCAA CTGAAACCGC GGGAGGAGGA AGCGCAGGAAT CAGGAACCTGG  
2001 CGGGGGTCCG CACCGGGCCT GAGTCGGTCC GAGGCCGTCC CAGGAGCAGC  
2051 TGCCCGTGCCT GGTACCTCTA GCCCCGGGGC CTGGAGGAGC GGTGGGAGCT  
2101 GGGGGCGCGG CAGGCAGGGG CAGAGCAGGC GTTCCGAGGG CCAGAGACCC  
2151 ACCCAGGTGG GGGTAGGGC CGCGGAAGGG CGGGGATGGC CGCAGGGGCA  
2201 GGGCTCAGGC TGTGGCGCC TGAGGCTTA GCTGGGGCAG GCTTGGCCTG  
2251 TCGAGGACCT GGGCAAGGGT GTCCCTGTAA GGGGTGGTGG GTGGAAGGGC

FIG. 3A

2301 CTGGGGAGGG AGGCTCCAGG TTGGCTCTG TTCCCAGC TGCGGAGGAG  
2351 ACCCTGACGC TAAGGAAGCA ATGAGGGCCA GTCCCCAGGC CAGGCTGCTG  
2401 CTGGGTACCC ATGGCTGCGT GTGAGCGAGG CAGGACCCCA CCTCCTCCCC  
2451 GTCTGCAGTC CATCCTGACC CTACAGTCCC CAGCCTCCTC GTCCCATGCC  
2501 TCCGTCTCCA GCTGCTGCCT TGCCCTCAGG GATGGCCCTT TTTCTGTCCC  
2551 CAGAACAGCA CTATGGGCTT CTCTTCTGAG CTGTGCAGCC CCCAGGGCCA  
2601 CGGGGTCTG CAGCAAATGC AGGAGGCCGA GCTTCGTCTA CTGGAGGGCA  
2651 TGAGAAAGTG GATGGCCAG CGGGTCAAGA GTGACAGGGGA GTATGCAGGA  
2701 CTGCTTCACC ACATGTCCCT GCAGGACAGT GGGGGCCAGA GCCGGGCCAT  
2751 CAGCCCTGAC AGCCCCATCA GTCAGGTGGG TCTCTATGGG ACTCTGGTGG  
2801 GTGCTGGCGT ATCTGCCCTC TCCTTCCTCT CCTGGGGGCC CTCTGGGGCA  
2851 GTGGCTGGAG ATCTGGCAGG CCAATGCTTG GGAGCCATTG TGCCCCCTC  
2901 CCTGCCTCCC CCATCTGTGC TGTATAGTCC TGGGCTGAGA TCACCAGCA  
2951 AACTGAGGGC CTGAGCCGCT TGCTGCGCA GCACGCAGAG GATCTGAAC  
3001 CAGGGCCCT GAGCAAGCTG AGCCTGCTCA TCCGGGAACG GCAGCAGCTT  
3051 CGCAAGACCT ACAGCGAGCA GTGGCAGCAG CTGCAGCAGG AGCTCACCAA  
3101 GGTGAGCGGG CAGCACTGGG GCTTCGGTCA TTTCTGTCTA AATTTTGAGC  
3151 CTCGAAGGGG TTGTTTGCA CAAGAGGCC TGGATTCACT GGGGAAGTGT  
3201 AAGTCCCTGA CCGCAGGCCT GGCTTGCTCT AACCTTGATG TAGCTTCCTC  
3251 TCTTCCTTCC CCTACGTTGA GCTGGCTTGC AGCAAGGCCT CTCTGTGCTT  
3301 TTTCTGTGCC TGGGCAAAGT GCTGGGAGTG TAAGGATGAG TGACCGGTCA  
3351 CGTGCCTGGG AGAACGTCAG AATCGGTACT CGCCTCCACA CTGTGCCATC  
3401 TGGCTCTGGG TTCTGAGAGT CAGGGAGAGG AATGAGGGTC AGTCTGTTG  
3451 CCTTCGACCT ATGCAGCCTC CTCTCAGGGC CCCAGAGACT GGGCAGCAGC  
3501 ATGGCCCCCCC GAAGGTCGAG GACTCGGCC GTGAAGTCAG CCTGCCTAGG  
3551 TTTGAATCCC ACCCAGCTCC TCAGTCTAGA GGCTGTGTGA TTTGGAAC  
3601 TTTATCTGGG AGCCTAGTGC CCCCATTCACT TGTGCTGGTC ACCCTCCCTG  
3651 CACCAACACCC CTTCCCTCAAG TGCAAGGCC AGCCTTGCCA TGGACCCACA  
3701 GCGGCCCTG GTGGCCACCC TGGCCCCATT CCTCGCCCCA AAAGATCATC  
3751 TGATTCAAGG GTGGGCCAT TTTTATAAAAG TTTTGCTGGA ACACAGCTAT  
3801 GCCCCTTGT TTTCATATTG TCTGTGACTA CAATGACAGA GTTGAGTAAT  
3851 TGTGACAGAG GCTCTATGGC CTACAAGCCT AAAATATTAA TTTACTATCT  
3901 GGCCCTTAA GAAAAGACT GATCTAGTCG AGGAATCTAG CTCAGTTACA  
3951 GATGGGGAAA CTGAGGTTGG GCGCTTGCCC AACATATCCC AGCACATAAA  
4001 CAGGAGAACT GGGACGAGAA CACTGATCTC GGGCTGTCT CATTTCTAC  
4051 TGCCAAGAAC ATAATTGCA GGACCCAGTG CAAAGTGAAA TTGTGGGGT  
4101 CTTTGTAAA AGATTGCTAG GAATTTCCAG GTGGAATAA TGGAGAATGA  
4151 AACCAAGCAC AGGGCCCTTC TACATGTGGA GCCCCGTGTG ACTGCACAGG  
4201 CCGTGCACAC CTGCAACTGG CCCTGCCTGC CACCAAGGCTA CCACTGTCAG  
4251 TCCAAGGAGG GACCGTTGTA GCCTGTAGTC TACCTCTTGT CCTCCCCAAG  
4301 GGGTCTGTCT TCAACAGGCT CTCTGATCTT TGACTCTCAC GTCAGCAGCC  
4351 AGCTTCTCA GAAGTCTCCA GGTGCTCCCT GCCTGACGAC AGGACCTTTC  
4401 CAGGGCTTCA CCCCCAGGCAA GAATCTTCCA CAACTGGGA CCTGCTGCC  
4451 CACACTGGCC TCTCCTCTCT CCCTAGACCC ACAGCCAGGA CATTGAGAAG  
4501 CTGAAGAGCC AGTACCGAGC TCTGGCACGG GACAGTGCCTC AAGCCAAGCG  
4551 CAAGTACCAAG GAGGCCAGCA AAGGTTGCTG GCTTCCCTTG CTGGCAGGGAA

4601 GGGAAATCCGA AGCCAGTGCT GACCTGTCCCT TGGGTACCCA GAGAGTGGGG  
4651 GCTGCCTGGG CCTCCATGCT GTCACTATA CCCCTGCCC CCCTTCTGGC  
4701 AGACAAGGAC CGTGACAAGG CCAAGGACAA GTATGTGCGC AGCCTGTGGA  
4751 AGCTCTTGC TCACACAAAC CGCTATGTGC TGGCGTGCAG GGCTGCGCAG  
4801 CTACACCACC AGCACCACCA CCAGCTCTG CTGCCCCGCC TGCTGCGTC  
4851 ACTGCAGGAC CTGCACGAGG AGATGGCTTG CATCCTGTAA GCCCGCAGCC  
4901 CCGTCCCCTG GCCCCCACCC TTGAGCAGCC CTAAGCCCAG CCATCAGGCC  
4951 CAGAGGCAGG ACCCAGAAAA TCCATTGCTG GGAAGGTGCT GGCCATGTAA  
5001 CCACATGAGA ACGGGACCTG GGCCAAGGAT TGGAAACAGG CAACTTACCT  
5051 CTGAATTACA CTATTCAGG GTCTCATTAT TCCAGGGTTT TATTACATT  
5101 ATTGAGCACT GTTCTGGCT CTGGATTATA CCAGAGAACG ATGGTAGACA  
5151 AAAACATCTG TCCTCAGGGA TCTTCGTGT TAGTGGAGTG AGAATGTGAG  
5201 GAGCACTAAG AGCCATGGAG AAAAATAAAG CAAGAGAAAGT GGATCGGGAC  
5251 CTGGGAGCAC GGAGGCAAGG GAGGAGGTGA CAGTTGTCCA TAGAGTGATC  
5301 TGAAAAGCC TCTTGAGAGG TGACATTCAA AGAGGCCCT GAGAGGGGTA  
5351 CGGGAGTGA TCATGGGCT ATTGGAGAA AGACCATTCC AGAAAGGAGG  
5401 ACAGCAATTAA CACAGGCCTT GAGGTAGGAG AGTACCAGGG ACTAATAGCC  
5451 AGGAACCAGT GGTGCCTCTG AGAGTGAGGG AGGGGGAGAG TCATACACGA  
5501 GGCTGGAGGA GGCAGGCCTC AAGGGCTACT GGGTGTAGA AGGTCTAGCA  
5551 GGGCCATGGT GAGGACTTTG GCTCTGGGTG AACAAAGAATG GCATGATCTG  
5601 ACCTCTGTTT TTTTGTTCAT TTTTGTTTA ACTTTTTTG AGTCAGAGTC  
5651 TCGCTCTGCC GCCCAGGCTA GAGTGCAGTG GCATGATCTC GGCTTACTGC  
5701 AACCTCCGCC TCCCAGGTTA AAGTGATTCC CCTGCCTCAG CCTCCCGAGT  
5751 AGCTGAAACT ACGGGCATGC GCCACCACAC CCAGCTAATT TTTGTATTT  
5801 TAGTAGAGAC GGGGTTTCAC CATGTTGCC AGGCTAGTCT CTAATTCTG  
5851 GGCTCAAAGC GATTTGCCTG CCTCTGCCTC CCAAAGTGCC GGGATTACAG  
5901 GCATGAGCCA CCATGCCAG CCCTGACCTC TGTTTTAATA AGGCCACTCT  
5951 GGCTGCTGTG CTGCAAATAG ACTTCAGGGA GCAAGGACAG AAGCTGGGAG  
6001 GCCAGAGAGC AGGCTGCTTGC CCATAATCCA GATCCAAGCT TTTGGCCAGC  
6051 TAGGACGGGG AGGTAGCAAT GGAGGTGAGG CGCGGTCAGG TCCTGGGGCA  
6101 GGTCTGGAA GGTGAAGCCA GTGGGATTTC CCTATGGATT GGAAGTGGGG  
6151 CGTAAATAG AGGAGTCAGG GGTCACTCTG GGGATTTGGC CTGGAGCAGC  
6201 TGGAAGATGG AGTGGCTGTT AACTTATGTA GGGAAAGGCTG TGGGAAGAAG  
6251 AGGTTAGGA GACAAGGATA GCAGTTCATT TATTTATTAA TTTATTATT  
6301 TATTTATTAA TTTATTAGA GATGTAGTCT CATTCTTCG CCAGGCTGGA  
6351 GTGCAGTGGC GCGATCTTGG CTCACTGCAA CCTCCACCTC CCAGGCTCAA  
6401 GCGATTCTCT TGCCTCAGCC TCCCAGTAG CCAAGTAGCT GGGACTACAG  
6451 GCATGTGCCA CCATGCCCTGG CTAATTTTG TATTTGCTTT TTCAGTAGAG  
6501 ATGGGGTTTC ACCACGTTAG CCAGGCTGGT CTCGAACTGA CCTCAGGCAA  
6551 TCCACCCGCC TCGACCTCCC AGTGTGGTA TTATAGGCGT GAGCCACTGT  
6601 GCCTGGCCCA CTGGATCCTT ATTACAACGT CCAGTGTCCC TCTTATATAT  
6651 ATCAGGAAAT AGAAGATTAG GGAGAGGTTA AATAATTGC CTAGAGTGGC  
6701 ATGGCTAGCT CGAAGTGAGG CAGGGGTCAA CCCCAGCCCT GACTCCAAAC  
6751 CCAGGGTCCT AGGCCTGAAC TGCCCAGCCT TGCCCAGCCT GAGGCTCCCC  
6801 TGACTGGGGA TCCCCTCTG GGGCAGGAA GGAGATCCTG CAGGAATACC  
6851 TGGAGATTAG CAGCCTGGTG CAGGATGAGG TGGTGGCCAT TCACCGGGAG

6901 ATGGCTGCAG CTGCTGCCG CATCCAGCCT GAGGCTGAGT ACCAAGGCTT  
6951 CCTGCGACAG TATGGTAAG CCCCCTCCTT GCTCCTGCTG GGCCCCAGGGC  
7001 TGCTGGCCTG TCCACTGACG GGGCGCTGTC CCCCACAGGT CCGCACCTGA  
7051 CGTCCCACCC TGTGTACAGT TCGATGAGTC ACTGCTTGAG GAGGGTGAAC  
7101 CGCTGGAGCC TGGGGAGCTC CAGCTGAACG AGCTGACTGT GGAGAGCGTG  
7151 CAGCACACGT GGGTGGTGGC TTTGCACCTG GGCTGCGGCG GGGCTCCCAG  
7201 CAGACCACGA GTGTTATGT AGGCAGGGCT AGGTGCTGGA GACTGTCCAC  
7251 ACAGAGCTGT CACCAGGTGG CCGGGCTTGC TTGGCTCTAC AGGGATGCAC  
7301 TGGACCTGGG TTGAGGGGGC AGGAGGGCTC GGTTCTAATG CTGCCCTTCT  
7351 CTTGGGTGCA GGCTGACCTC AGTGACAGAT GAGCTGGCTG TGGCCACCGA  
7401 GATGGTGTTC AGGCAGCAGG AGATGGTTAC GCAGCTGCAA CAGGAGCTCC  
7451 GGAATGAAGA GGAGAACACC CACCCCCGGG AGCGGTGAGT GGGCCCCCTGC  
7501 CTGCAGCAGC CTCCTGGCC TCCCTCCCTC CTACCTACCC TAACTGCTGC  
7551 TGGCTAGCCG CCGCAGACCG AGCCCTTATT CTTCATCCAC CCTCCCACCC  
7601 GCCCCTGCCT GCAGGGTGCA GCTGCTGGC AAGAGGCAAG TGCTGCAAGA  
7651 AGCACTGCAG GGGCTGCAGG TAGCGCTGTG CAGCCAGGCC AAGCTGCAGG  
7701 CCCAGCAGGA GTTGTGCAG ACCAAGCTGG AGCACCTGGG CCCCAGGAG  
7751 CCCCCGCCTG TGCTGCTCCT GCAGGATGAC CGCCACTCCA CGTCGTCCCTC  
7801 GGTGAGCTGC CCCATCCGCG GCCGCTGCC CCAACCGGCC TGCCCACCTG  
7851 GGGCTGCGCT CCTCATTTTC GCCCTCCCCC TCCCTAAAGCC TGGCCACCCG  
7901 CTGACGTCTG TCCCTGGCCT CAGGAGCAGG AGCGAGAGGG GGGAGGACA  
7951 CCCACGCTGG AGATCCCTAA GAGCCACATC TCAGGAATCT TCCGCCCCAA  
8001 GTTCTCGGTG AGTGGCGCCC AGCCTGGGCC CCCCTACTGT TGTGTTTCGA  
8051 GTTTAATCAC TGGGATGTCC TAGAGAGGAG GCTCTGCCA GGCTGCTTGT  
8101 ATTGGGAAGT TCCTCTCTTC CCTGGGATTC CAGGCTGCAG ATGTCCCCAG  
8151 ACCCTGCCCC TGTGACCCCT CCCTTCCAT CGCCCCAGTG TGCTAAAGGG  
8201 ACCAGCAACC TCGACTATTG CATGGCTCTC CCTGCTTCAG GAGCGGTTGG  
8251 GGGCCTGTGG CCTGGAGGAG GAGGCACCAAG CTTGGTTTGG GGTCTTCCTG  
8301 CCTGGGCTTC CCTTCCCAGC TCTGCCAGC GTGAGCCTGG GCCAGTCCAG  
8351 TGCCCCTCTC AGGGGCCTGT GGATGGCTCT GCATGCCACT CCATGGTTGT  
8401 AAGGGCTGAG GGCATATAGG GGGGAGAGAG AGACCCCCGG CTGCCACAC  
8451 GGCCTCTCA ACAAGGTGGT TAAGTGACTC CCTCTCGATC CTCCCTGCC  
8501 CAGCTCCCTC CACCGCTGCA GCTCATTCCG GAGGTGCGAG AGCCCCCTGCA  
8551 TGAGCAGCTG TGGTACCAAG GGGCCATCCC GAGGGCAGAG GTGGCTGAGC  
8601 TGCTGGTGCA CTCTGGGAC TTCCCTGGTC GGGAGAGCCA GGGCAAGCAG  
8651 GAGTACGTGC TGTCGGTGCT GTGGGATGGT CTGCCCTGGC ACTTCATCAT  
8701 CCAGTCCTTG GATGTGAGTG GGGCTGGAC CCGAGCCTTC CAGGCCTCAC  
8751 TCTTCCCTTC CCTTCCCCTC CCCAAGGGAA ATGGCTTTC AGGGTAGGGG  
8801 GTAGCTGCCA GGTCTTGGAT GCCTCCCTAG CAGGGCTGGC TGGAAGGGC  
8851 CACAGAGACC ACCCTGTCCC TGCAACAAAA TAGAGGCTTA AGTGTGAGTC  
8901 CTCCCCTGGT GGGCAGCAG GATGTGATGT GCCATCAGAT GGCATCTTT  
8951 CTGGAGGTCT CTCTGCCCT GGTCCCTGGC AGGCCCTTC TCCCCTGCTG  
9001 CTCTCCCTTT CCCCCCTCCA GGGCTCACGC CCCCTCAGAA TGGAGGCTGC  
9051 TGACCCCGGG TCCCCCTGCC TGCAAGACCT GTACCGACTG GAAGGGAG  
9101 GCTTTCTAG CATTCTTTG CTCATCGACC ACCTACTGAG CACCCAGCAG  
9151 CCCCTCACCA AGAAGAGTGG TGTGTCCTG CACAGGGCTG TGCCCAAGGT

FIG. 3D

9201 GAGCCTGCAC CCAGCCTGGC CCATGCCACC TGTGGCAGGG CTTGGGGAGT  
9251 GTGGGTCAAG CCCACCCAGC GTCTGAGCAG AAAGGGCTTT CCAGGCCCTC  
9301 CGTCTACATA CAAGATGCAG AGTGAGTGAC CCTCAGGGCC AGCCTTGCTC  
9351 TAGGTTTGGA ATGTCAGGGC CACTCCTATG CCATGGGCTG TACACACCA  
9401 GTTGGTGCTT ACCTGGTCAG GGCACCTGCC TGGACCCCGT AGTCATCTCA  
9451 GTGTGCTCCC CACGTGGTCC CACCCCTGGT CACATATGGA GGCGCCAAAA  
9501 AATGGAGGAC ACAGCCCTTC TAAGGGCCA GCACCCCTTT TCTTCAGACT  
9551 TCTGATCCCC TGTCTCCTCT CTTCCCCAGG ACAAGTGGGT GCTGAACCAT  
9601 GAGGACCTGG TGTTGGGTGA GCAGATTGGA CGGGTGAGTG CGCCTCTGCT  
9651 GGCCTCCTTG TCGCTGGCGA CTTCTCCTGA GTCCGGCCTG GGCCCCCTGC  
9701 CCTACCACCC AGAAACCTCC CTGCCCCATC TGATTCCCCA CTTGTACCCC  
9751 GACTCCCTGC CCAGCCCCA CCACACACCA TCCTCCAGGA AACGGGACAG  
9801 TACCTACGCT GAAAACCTCC AGCAGACAGC TCTGCCAGCA CCCTGACCTC  
9851 ATCACCCCCA CCCAGGCCGC CCCCATCGAG CTCTTGTGTG CACGCAGGG  
9901 GACACCCCTGT TACTGTAAGC CATAAGATAC CTGTTAGGG AAGAAGTCAC  
9951 TGTCTAAAAA ATCAGAATGC TTTTCAAACC CAAGGGAGAG TGATTTTGG  
10001 ATTTCCATGT CACTTCTCTC AGGAAGGGTG GCACATCGGA GGCAACTTTC  
10051 CCTGCTGCCCCC CCATGTGCTC TCTAGGTTCC CCAGCGAGGG TCAAACCTCC  
10101 AGAGAGCCTG GGTGGAGGGG TCCGAACACG GGGGCCCCCTC ACCCAGGGT  
10151 AGGAAGCAGA ATGGGTAGGA AGCGGAGAAG AGAACTGCGG GACTGGGAAG  
10201 GCCGTGGTAG GAGCCAAGA CCGTTTCAGG GGAACCTTGG CGAAGTGTTC  
10251 AGCGGACGCC TGCGAGCCGA CAACACCCCTG GTGGCGGTGA AGTCTTGTG  
10301 AGAGACGCTC CCACCTGACC TCAAGGCCAA GTTTCTACAG GAAGCGAGGT  
10351 GGGTGATAAA CTAATGATCA CCACGGGTCC CGCATAACACA GAGGTTACAC  
10401 TGCATGGCAC AGTGTGAAGT GCTTGACAC CGTGGTGGTG TTTAGTCTC  
10451 GAGGCCCCCCC ATTGCGGGTA GTACCCCTT ATAGTGCCGA AGGGTAGAGG  
10501 CTGCCCCAGG TCACACGTCC GGGTCTGCTG GCCTTGGAGG CCAAGCTCTT  
10551 CTCCCATCAT CCCTGGGGGG CCCTGGGGAG GCGGGCCTGG CCACGTAGAT  
10601 CCTGAGCAGC AGTGCCCTCC AGGATCCTGA AGCAGTACAG CCACCCCAAC  
10651 ATCGTGCCTC TCATTGGTGT CTGCACCCAG AAGCAGGCCA TCTACATCGT  
10701 CATGGAGCTT GTGCAGGGTG AGCGCGGGGC GCTGAGCTCC AGGTAGGGCG  
10751 CGCAGCCTGG TCAGGTGGCA GCCTTACCTC AGGAGGCTCA GCAGGGTCC  
10801 TCCCCACCTG CAGGGGGCGA CTTCTGACC TTCCTCCGCA CGGAGGGGGC  
10851 CCGCCTGCCGG GTGAAGACTC TGCTGCAGAT GGTGGGGGAT GCAGCTGCTG  
10901 GCATGGAGTA CCTGGAGAGC AAGTGCTGCA TCCACCGGTG AGTGGCGGT  
10951 GGCCACGGGC CCTGCCAACA CCCCCGACCA GAGTCAGAG GTACCTATAC  
11001 CCCTAGGGCC CCCCCGTGGA CCATCAGGCA TCAGCTCCAG AGGGGGAGTT  
11051 GGCCTCTGTG GTAGACAGGG GTGCCAGGG CCGGGAGCAG CTTTTGTCTC  
11101 TGGCTTCCT AGAGTGTCA GCCAGGGCTG GGCAGGCGAC TGTTGGCCAA  
11151 ATGAGCCCCC GCCCTGTCTC ACCCAGGGAC CTGGCTGCTC GGAACCTGCCT  
11201 GGTGACAGAG AAGAATGTCC TGAAGATCAG TGACTTTGGG ATGTCCCGAG  
11251 AGGAAGGCCA TGGGGTCTAT GCAGCCTCAG GGGGCCTCAG ACAAGTCCCC  
11301 GTGAAGTGGA CCGCACCTGA GGCCCTTAAC TACGGTACCT AGTCCCTGTC  
11351 TACCCCTGGAC TCCATGGCCA GAGGCCAGGC CTGGGTCTG CCGGCTGCCT  
11401 CGCCCTGGCC CCAGGGAGGG TGCACTCACG CTGCCTCACC TCCTCGCTC  
11451 CTCTGCAGGC CGCTACTCCT CCGAAAGCGA CGTGTGGAGC TTTGGCATCT

11501 TGCTCTGGGA GACCTTCAGC CTGGGGCCT CCCCTATCC CAACCTCAGC  
11551 AATCAGCAGA CACGGGAGTT TGTGGAGAAG GGTAAGCACC CTGTGATGAC  
11601 AGCAGCCTCA GGCTGCACCC TCTTCCAGAT GCTCCAGCCG GACTCTTCTA  
11651 ACTCCCTTAA TGCCAACCTT CCCACCAGGC AGAATAAGAA TAACCTGGCC  
11701 AGTTGTCAC GCCTGTCATC CCAGCACTT GGGAGGCTGA GCTGGGTGGA  
11751 TCACTTGAGC CCAGGAGTTC AAGATCAGCT TGGACAACAC AGTGAACACTC  
11801 CATCTGTACA AAAAATACAA AAATAGACTG GGCACGGTGG CTCACACCTG  
11851 TAATCCCAGC ACTTTGGGAG GCCGAGGCAG GTGGATCACC TGTGGTCAGG  
11901 AGTTTGAGAC CAGCCAGACC AACATGGTGA AACCCCATCT CTACTAAAAA  
11951 TACAAAAATT AGCCAGGCAT GGTGGCACGT GCCTGTAATC CCAGCTACTT  
12001 GGGAGGCTGA GGTGGGAGAA TTGCTTGAAC CCAGGAGGCG GAGGCTGCAG  
12051 TGAGCCGAGA TTGTGCCACT GCACTCCAGC CTGGGCGACA AGAGTGAAC  
12101 TCCATCTCAA AAAAACCACAA AAAACAAAAAA ATACAAAAAT TAGCTGGTG  
12151 TGGTGACATG CGCTGTAGT CCCTGCTACT CGGGAGGCTG AGGTGGGAGG  
12201 ATCACTGGAG CCCGGGAGGT GGAGGTTGCA GTGAGCTGAG ATCATGCCAC  
12251 TGCACCCCAA CCTGGGTGAC AGAGAGAGAG AGAGACCTTG ACTCGAAAAA  
12301 GAAAAAAACC TGGGCGCAGT GGCTCACGCC TGTAATTCA ACATTTGGG  
12351 AGGCTGAGGA AGGTGGATCA CTTGAGTCTA GGAGTTGAC ACTAGCCTGG  
12401 CCAACATGGC AAAACCTGTC TCTACTAAAA ATACAAAAAA TTAGCGAGGT  
12451 GTAGTGGTGC AAGCCTGTAA TCCCAGCTAC TTGGGAGGCT GAGGCACAAG  
12501 AATCGCTTGA ACCTGGGAGG TGGAGGTTGC AGTGAGCTGA GATCACACCA  
12551 CTGCATTCCA GCGTGGGTGA CAGAGCAAGA CTCCATCTCA GAAAAAGAAA  
12601 AAAAATAAATGAAATATCCCT GTAGCTACTA CTGAGTGAGC ACCTGGTCTG  
12651 TGCTAGGTCA CATGTTATTCTT CATTGCTCA TCACTACATG TGTGGTAGGG  
12701 ATTAATATGT CCCTTCTCA GATGGAAAAA CAGGCTGGCA GAGGGGACAC  
12751 AGCTAGCAGG TGGTAGGATT AGGATCAGAA GCCAGGCCTC TTTGTCCCTT  
12801 GGGCCCTTGG TGGAGAACAG TGCATCCTTC AGAACAGTGC ATCTTAAGCA  
12851 GCTCCTATGG CTCATGGTAT CCCCCAGAGT CTGCCGAGGA CCCTCAAAC  
12901 CCCTCCTCAT GCCTGGTGTG CTGTGCCTCT CCTCACAGGG GGCGTCTGC  
12951 CCTGCCAGA GCTGTGCTCT GATGCCGTGT TCAGGCTCAT GGAGCAGTGC  
13001 TGGGCTATG AGCCTGGCA GCGGCCAGC TTCAGCACCA TCTACCAAGGA  
13051 GCTGCAGAGC ATCCGAAAGC GGATCGGTG AGGCTGGGAC CCCCTCTCA  
13101 AGCTGGTGGC CTCTGCAGGC CTAGGTGCAG CTCCCTAGCG GCTCCAGCTC  
13151 ATATGCTGAC AGCTCTTCAC AGTCCCTGGAC TCCTGCCACC AGCATCCACA  
13201 CTGCCGGCAG GATGCAGCGC CGTGTCCCTCT CTGTGTCCCT GCTGCTGCCA  
13251 GGGCTTCTC TTCCGGGAG AAACAATAAA ACCACTTGTG CCCACTGAAC  
13301 ACTCCTGGCA TGTGCACTCC TCTGGAGGC AGGTCTCAGA AGGCACAAGT  
13351 GCCGGTATGG TGGCCTTGGG GAAGGAGGAG GACAGGCAGT ATGCATGGGG  
13401 CAGAGCTGAC ATGATTAGT AGCAGCTGGA TGTGAGACAT GCGGAAGGCG  
13451 GGGGAGAGAT CAGGATGATA TACAGGCTAT GGCCAGATGG CGGTGTCTAC  
13501 CCCTGAAATA GGATTATAGG AAGAGGATCA GAGCTTCGAG GAGGATGTTG  
13551 AGTTTAGAGA TGTTGCATT TATTGGAGAT AAAAGTGTGG GTGAAGCCAG  
13601 GTGTGGTGGT AGACACCTGT AGTCCCAAGGT ACTTGGGAGG CCAAGGCATG  
13651 TGGATTGCTT GAGCCTAGTT TGAGACCAGC CTGGGCAACA TGGCAAAACT  
13701 CCATCTTAC AAAAACCACAA AACAAAAAAC AAAAACCACAA GTAAAATTAG  
13751 CCAGGCGTGG TGGCACACAC CTATAGTCCC AGCTACTCAG AAGGCTGAGG

13801 TAGGAGGATC AATTGAGCCT CGGAGGTCGA GGCTGCAGTG AGCTGTGATC  
13851 ACACCACTGC ATTCCAGCCT GGGCAACAAA GCGAGGCCCT GTCTCAAAAA  
13901 TAAGTAAATA AAAATAATAA ATAATTAAATT TAAAATGTAG ATGAATAGGT  
13951 CTGGAAGCCC AGATGGAGAT GAAGGCTGGC AATAGATGTG TGAATCATTG  
14001 GCTTATGAAT ATTAGAGAGT AGCTGACACT ATGGATGCGT ATAACACTCG  
14051 CATAAAATTC AGGAGGAGAT GAGAAGAGAG TTCCACTCAA AGAAGACTGA  
14101 TGTGGCTGAT GAGGAAGAAA ATGCCTTGA GGGAGTTGTT TCTCAAGATG  
14151 AATTATTGA GGAATAAGAT GGCAGACTGG GGAGCCTTCA CCTCCTCCCC  
14201 TAAGTCCCAG TGAAACCTAA AAAGTCATCT GAAATATTAA CATCACCAAA  
14251 AGCGAAGTTT GAGAAGATAA GGAAGTATGA ACATAACTAA AAAACAAAGT  
14301 GGGAAACATT TGTAATACAG AACAGGGCAA TGAAAACCTT GAAGTAAAAT  
14351 GCCCATCCCT CAAGAAAGTT CAGGAAATAG TTAACATCAG CTGGGTGCAG  
14401 TGCTCACAC CTATAATCCC AGCACTTGG AAGGCTGAGG CAGGTGGATC  
14451 ACCTGAGGTC AGGAGCTGA GACCAGTCTG GCCAACATAG TGAAACTCCG  
14501 TCTCTGCTAA AAATACAAAA AAAATTAGCC AGGCGTGGTG GTGTGCACCT  
14551 GTAATCCCAG CTACTCTGGA GGCTGAGAAG GGAGAATTCC TTGAACCGGG  
14601 GAGATGAAGG TTGGAGTGAG CAGAGACCGC GCCATTGCAC TCCAGCCTGG  
14651 GCAACAAGAG CGAAGAACAA AACTATGTCT CAAAAAAACA AAACACAGCA  
14701 AACAAAAATC TATTTGAAA GAGATGAGAG TGAGCCATAT AACTTGTITA  
14751 AACAAAAGGA AGTTGTGTTG TCGTGTAAATT AAATGAAAAT ACTAGGAAGT  
14801 GAAATAATAC CTCCAATGGA AATGGTAGAA AGCAGAACTG AAAAACTTCT  
14851 GCTAGGTAGG ATATGGTAGG TCTCTGCACG CCACCACTCC CATTGCAACC  
14901 GCTAGGGAAA AAACAGCTAA GATGAAAATG TCTTTTTTT TCTTTTTTT  
14951 TTTTTTTGATGGAGTCT CGCGCTGTTG CACAGGCTGG AGTGCAGTGG  
15001 CGCGATCTCA GTTCACTGCA ACCTCTGCCT CTCGGGTTCA AGCGATTCTC  
15051 CTGCCTCAGC CTCCTGAGTA GCTGGGATTA CAGGCACGCA TCACTCACGA  
15101 GCGGCTAATT TTTGTAATTT TAGTAGAGAC GGGGTTTCAA CATGTTGGTC  
15151 AGGCTGGTCT CAAACTCCTG ACCTCAAAGT GACCCGCCA CCTCGGCCCTC  
15201 CCAAAGTGTGTT GGGATTACAG GGATGAGCCA CCACGCCTGG CCGAAATGTC  
15251 TTATTTTAA AAAGAATGAA GAGTGGTCAC AGAAATAAAG ACTGAAT (SEQ ID NO:3)

FEATURES:

Start: 2563  
Exon: 2563-2775  
Intron: 2776-2927  
Exon: 2928-3101  
Intron: 3102-4476  
Exon: 4477-4573  
Intron: 4574-4702  
Exon: 4703-4886  
Intron: 4887-6827  
Exon: 6828-6965  
Intron: 6966-7038  
Exon: 7039-7158  
Intron: 7159-7361  
Exon: 7362-7484

FIG. 3G

Intron: 7485-7614  
Exon: 7615-7801  
Intron: 7802-7923  
Exon: 7924-8007  
Intron: 8008-9075  
Exon: 9076-9198  
Intron: 9199-9579  
Exon: 9580-9633  
Intron: 9634-10229  
Exon: 10230-10348  
Intron: 10349-10622  
Exon: 10623-10717  
Intron: 10718-10813  
Exon: 10814-10937  
Intron: 10938-11176  
Exon: 11177-11334  
Intron: 11335-11458  
Exon: 11459-11581  
Intron: 11582-12938  
Exon: 12939-13078  
Stop: 13079

CHROMOSOME MAP POSITION:

Chromosome 15

ALLELIC VARIANTS (SNPs):

| DNA      |       |       |                | Protein  |       |       |
|----------|-------|-------|----------------|----------|-------|-------|
| Position | Major | Minor | Domain         | Position | Major | Minor |
| 28       | C     | T     | Beyond ORF(5') |          |       |       |
| 847      | A     | G     | Beyond ORF(5') |          |       |       |
| 2159     | G     | C     | Beyond ORF(5') |          |       |       |
| 2484     | C     | T     | Beyond ORF(5') |          |       |       |
| 2577     | T     | C     | Exon           | 5        | S     | S     |
| 2808     | -     | T C   | Intron         |          |       |       |
| 2922     | G     | A     | Intron         |          |       |       |
| 4312     | C     | T     | Intron         |          |       |       |
| 4903     | G     | A     | Intron         |          |       |       |
| 5193     | A     | G     | Intron         |          |       |       |

Context:

DNA  
Position

|    |   |
|----|---|
| 28 | CTGGCCACCAGGCTGGCGAGCCAAGG<br>[C,T]<br>CGAAGCTCTGGCTGAACCCTGTGCTGGTCTGACCACCCCTCCCTCTCTGCACCCGC |
|----|---|

FIG. 3H

CTCTCCCGTCAGGGCCAAGTCCCTGTTTCTGAGCCGGCTGCCCTGGCTTTGGCAC  
 TCACAGACCTGGAGCCCTGGGTGGTGGGGAGGGCGCTGGCCAGCCGGCTCTC  
 TGGCCTCCACCCGATGCTGCTTCCCTGTGGGGATCTCAGGGCTGTTGAGGATATA  
 TTTTCACTTTGTGATTATTCACTTAGATGCTGATGATTGTTTGTATTTAATGG

847 ATGCCGGGGTTCATAGGTCACTGGCTCTCAAGTGCCAGAGGTGGCAGGTGGCAC  
 TGAGCCCCCCCACACTGTGCCCTGGTGGAGAAAGCACTGACCTGTCATGCCCTCAA  
 ACCTCTCTCTGACGTGCCCTTGCACCCCTCCATTAGGACAATCAGTCCCTCCAT  
 CTGGGAGTCCCCCTTCTTCTACCCCTAGCCATTCTGTACCCAGCCATGCCAGG  
 GGTGCCCTCTCCATCCCCCTGCCCTCGTGGCAGCCGGCTGGTTTGTAAAGAT  
 [A, G]  
 CTGGGTGGTGCACAGTGATTTTCTTGTAAATTAAACAGGCCAGCATTGCTGGTTC  
 TATTAATGGACATGAGATAATGTTAGAGGTTAAAGTGATTAACGTGCAGACTATGC  
 AAACCAGGCCAGTCTCCAGTGTGGTACCGTTCTGCATCGCAGCTGAGGATAGGGG  
 GCCAGTTAGGCCTACACAGTGGCTGCCTGCCTGGATGTGGCCAAAGTCAGAAGGCCAA  
 AGTCCTCCAAGGGCGGGAGGATGCCAGCCCTAGTGGAGGAGCTGGTCCCCCTGGGG

2159 TGCGGGCCAGCGGGAGCAGGGGGCTGGCAGGGCTGGCAGGTGGCCGGTCCGCC  
 CGCCCCCTCCCTCCACAGGCCGCCGGGCTGGCCAAGTGAACCGCGGGAGGAG  
 GAAGCGCGGAATCAGGAACCTGGCGGGGTCGCACCGGGCTGAGTCGGTCCGAGGCCGT  
 CCCAGGAGCAGCTGCCGTGCGGGTACCTCTAGCCCCGGGCTGGAGGAGCAGTGGAG  
 CTGGGGCGCCAGGCAGGGCAGAGCAGGCGTCCGAGGGCAGAGAACCCACCCAGGT  
 [G, C]  
 GGGGTAGGGGCCCGGAAGGGCGGGATGGCCGCAGGGCAGGGCTCAGGCTGTGGCGC  
 CTGAGGCTTCAGCTGGGGCAGGCTGGCTGTCAGGACCTGGCAAGGGTGTCCCTGTA  
 AGGGGTGGTGGGTGGAAGGGCCTGGGGAGGAGGCTCAGGTTGGCTCTGTTCCGAAC  
 GTGCGGAGGAGACCTGACGTAAGGAAGCAATGAGGGCAGTCCCCAGGCCAGGCTGCT  
 GCTGGGTACCCATGGCTGCGTGTGAGCGAGGAGGCCACCTCCCTCCCGTCTGCAGT

2484 GGATGGCCGCAGGGCAGGGCTCAGGCTGTGGCGCTGAGGCTTCAGCTGGGCAGGCT  
 TGGCCTGTCGAGGACCTGGCAAGGGTGTCCCTGTAAGGGGTGGTGGTGAAGGGCCTG  
 GGGAGGGAGGCTCAGGTTGGCTCTGTTCCGAACGTGCGGAGGAGACCTGACGCTAA  
 GGAAGCAATGAGGGCAGTCCCCAGGCCAGGCTGCTGGTACCCATGGCTGCGTGTG  
 AGCGAGGCAGGACCCACCTCCCTCCCGTGCAGTCCATCCTGACCCCTACAGTCCCCAG  
 [C, T]  
 CTCCCTGCTCCATGCCCTCGTCTCCAGCTGCTGCCCTCCAGGGATGGCCCTTTCT  
 TGTCCCCAGAACAGCACTATGGGCTCTTCTGAGCTGTGAGGCCAGGGCCACGGG  
 GTCCTGCAGCAAATGCAGGAGGCCAGCTGGCTACTGGAGGGCATGAGAAAGTGGATG  
 GCCCAGCGGGTCAAGAGTGACAGGGAGTATGCAGGACTGCTTACCCATGTCCCTGAG  
 GACAGTGGGGCAGAGGCCAGGCCACCTCCCTCCCGTGCAGCAGGCCATCAGTCAGGTGGTCTC

2577 GTAAGGGTGGTGGGTGGAAGGGCCTGGGGAGGGAGGCTCAGGTTGGCTCTGTTCCCG  
 AACGTGCGGAGGAGACCTGACGCTAAGGAAGCAATGAGGGCAGTCCCCAGGCCAGGCT  
 GCTGCTGGTACCCATGGCTGCGTGTGAGCGAGGAGGCCACCTCCCTCCCGTCTGC  
 AGTCCATCCTGACCCCTACAGTCCCCAGCCTCCCTGACAGGCCATCAGTCAGGTGGTCTC  
 GCCTGCCTCCAGGGATGGCCCTTTCTGCCCCAGAACAGCACTATGGGCTCTCTTC

|      |        |   |
|------|--------|---|
|      | [T, C] | GAGCTGTGCAGCCCCCAGGGCACGGGTCTGCAGCAAATGCAGGAGGCCAGCTCGT<br>CTACTGGAGGGCATGAGAAAGTGGATGGCCAGCGGGTCAAGAGTGCAGGGAGTATGCA<br>GGACTGCTCACACATGTCCTGCAGGACAGTGGGGCAGAGCCGGCATCAGCCCT<br>GACAGCCCCATCAGTCAGGTGGTCTATGGACTCTGGTGGTGTGGTATCTGCC<br>TTCTCCTCCTCTGGGGCCCTGGGCAGTGGCTGGAGATCTGGCAGGCCAATGC  |
| 2808 |        | CCAGCTGCTGCCTTGCCTCCAGGGATGGCCCTTTCTGCCCCAGAACAGCACTATGGG<br>CTTCTCTCTGAGCTGTGCAGCCCCCAGGGCACGGGTCTGCAGCAAATGCAGGAGGC<br>CGAGCTTCGTCTACTGGAGGGCATGAGAAAGTGGATGGCCAGCGGGTCAAGAGTGCAG<br>GGAGTATGCAGGACTGCTCACACATGTCCTGCAGGACAGTGGGGCAGAGCCGGGC<br>CATCAGCCCTGACAGCCCCATCAGTCAGGTGGTCTATGGACTCTGGTGGTGTGG<br>[-, T, C]<br>GTATCTGCCTTCTCCTCTGGGGCCCTGGGCAGTGGCTGGAGATCTGGCA<br>GGCCAATGCTTGGGAGCCATTGTGCCCTCTGGCTCCCCATCTGTGCTGTATAGT<br>CCTGGCTGAGATCACCAGCAAACGTAGGGCTGAGCCGCTGCTGCCAGCACGAG<br>AGGATCTGAACTCAGGGCCCTGAGCAAGCTGAGCCTGCTATCCGGAACGGCAGCAGC<br>TTCGCAAGACCTACAGCGAGCAGTGGCAGCAGCTGCAGCAGGAGTCACCAAGGTGAGCG     |
| 2922 |        | GGAGGCCAGCTCGTCTACTGGAGGGCATGAGAAAGTGGATGGCCAGCGGGTCAAGAG<br>TGACAGGGAGTATGCAGGACTGCTCACACATGTCCTGCAGGACAGTGGGGCAGAG<br>CCGGGCCATCAGCCCTGACAGCCCCATCAGTCAGGTGGTCTATGGACTCTGGTGG<br>TGCTGGCGTATCTGCCTTCTCCTCTGGGGCCCTCTGGGCAGTGGCTGGAGA<br>TCTGGCAGGCCATGCTTGGGAGCCATTGTGCCCTCTGGCTCCCCATCTGTGCT<br>[G, A]<br>TATAGTCTGGCTGAGATCACCAGCAAACGTAGGGCTGAGCCGCTGCTGCCAGC<br>ACGCAGAGGATCTGAACTCAGGGCCCTGAGCAAGCTGAGCCTGCTCATCCGGAACGGC<br>AGCAGCTCGCAAGACCTACAGCGAGCAGTGGCAGCAGCTGCAAGCAGGAGTCACCAAGG<br>TGAGCGGGCAGCACTGGGCTCGGTCTTGTCTAAATTGAGCCTCGAAGGGGTT<br>GTTTGACAAGAGGCCCTGGATTCACTGGGAAGTGTAACTCCCTGACCGCAGGCCCTGG         |
| 4312 |        | GGACGAGAACACTGATCTGGCTGTCATCTATTCTACTGCCAAGAACATAATTGAG<br>GACCCAGTGCAAAGTGAATTGTTGGGGTCTTGTAAAGATTGCTAGGAATTCCAGG<br>TGGCAATAATGGAGAATGAAACCAAGCACAGGGCCCTCTACATGAGGCCCCGTGTGA<br>CTGCACAGGCCGTGACACCTGCAACTGGCCCTGCCACCAGGCTACCAACTGTCAGT<br>CCAAGGAGGGACCGTTGAGCCTGAGTCTACCTCTTGCCTCCCCAAGGGTCTGTCTT<br>[C, T]<br>AACAGGCTCTGATCTTGAACGTCAGCAGCCAGCTTCCCAGAACAGTCTCCAGG<br>TGCTCTTGCCTGACGACAGGACCTTCCAGGGCTTACCCAGGCAAGAACATCTCCACA<br>ACTGGGACCTGCTGCCAACACTGGCTCTCCCTCTAGACCCACAGGCCAGGACA<br>TTGAGAAGCTGAAGAGCCAGTACCGAGCTGAGCAGTGGCACGGACAGTGCCAAGCGCA<br>AGTACCAAGGAGGCCAGCAAAGGTTGTCGGCTCCCTGCTGGCAGGGAGGAAATCCGAAG |
| 4903 |        | GAATCCGAAGCCAGTGTGACCTGTCCTGGTACCCAGAGAGTGGGGCTGCCCTGGGC<br>TCCATGCTGTCATCTATACCCCTGCCCTCTGGCAGACAAGGACCGTGACAAGGCC<br>AAGGACAAGTATGTGCGCAGCCTGTTGAAGCTTGTCTTGTCTACCACAAACCGCTATGTGCTG  |

GGCGTGGGGCTGCGCAGCTACACCAACCAAGCACCACCAAGCTCCTGCTGCCGGCTG  
CTGCGGTCACTGCAGGACCTGCACGAGGAGATGGCTTGATCCTGTAAGCCCGAGCCCC  
[G,A]  
TCCCCTGGCCCCACCCCTGAGCAGCCCTAACGCCAGCCATCAGGCCAGAGGCAGGACC  
CAGAAAATCCATTGCTGGAAAGGTGCTGGCCATGTAACCACATGAGAACGGGACCTGGGC  
CAAGGATTGGAAACAGGCAACTTACCTCTGAATTACACTATTCCAGGGTCTCATTATTCC  
AGGGTTTATTACATTGAGCACTGTTCTGGCTCTGGATTATACCAGAGAACGATG  
GTAGACAAAAACATCTGTCCTCAGGGATCTTCGTGTTAGTGGAGTGAGAATGTGAGGAG

5193 CCGCAGCCCCGTCCCCCTGGCCCCACCCCTGAGCAGCCCTAACGCCAGCCATCAGGCCA  
GAGGCAGGACCCAGAAAATCCATTGCTGGAAAGGTGCTGGCCATGTAACCACATGAGAAC  
GGGACCTGGGCCAAGGATTGGAAACAGGCAACTTACCTCTGAATTACACTATTCCAGGGT  
CTCATTATTCCAGGGTTTATTACATTGAGCACTGTTCTGGCTCTGGATTATACC  
AGAGAACGATGGTAGACAAAAACATCTGTCCTCAGGGATCTTCGTGTTAGTGGAGTGAG  
[A,G]  
ATGTGAGGAGGACTAAGAGCCATGGAGAAAAATAAGCAAGAGAAGTGGATCGGGACCTG  
GGAGCACGGAGGCAAGGGAGGAGGTGACAGTTGTCATAGGTGATCTGGAAAGCCTCT  
TGAGAGGTGACATTCAAAGAGGCCCTGAGAGGGGTACGGGAGTGAATCATGGGCTATT  
TGGAGAAAAGACCATTCCAGAAAGGAGGACAGCAATTACACAGGCCTTGAGGTAGGAGAGT  
ACCAGGGACTAATAGCCAGGAACCAGTGGTGCCTCTGAGAGTGAGGGAGGGGAGAGTC